

# FIRST ORDER PROBABILITIES FOR GALTON-WATSON TREES

MOUMANTI PODDER AND JOEL SPENCER

**ABSTRACT.** *In the regime of Galton-Watson trees, first order logic statements are roughly equivalent to examining the presence of specific finite subtrees. We consider the space of all trees with Poisson offspring distribution and show that such finite subtrees will be almost surely present when the tree is infinite. Introducing the notion of universal trees, we show that all first order sentences of quantifier depth  $k$  depend only on local neighbourhoods of the root of sufficiently large radius depending on  $k$ . We compute the probabilities of these neighbourhoods conditioned on the tree being infinite. We give an almost sure theory for infinite trees.*

## 1. INTRODUCTION AND MAIN RESULTS

For  $\lambda > 0$  we let  $T_\lambda$  denote the standard Galton-Watson tree, in which each node independently has *Poisson* offspring with mean  $\lambda$ . We shall set

$$p = p(\lambda) = \Pr[T_\lambda \text{ is infinite}]. \quad (1.1)$$

As is well known, when  $\lambda \leq 1$ ,  $p(\lambda) = 0$  while when  $\lambda > 1$ ,  $p$  is the unique positive solution to the equation

$$1 - p = e^{-p\lambda}. \quad (1.2)$$

We let  $T_\lambda^*$  denote  $T_\lambda$  conditioned on  $T_\lambda$  being infinite. (When using  $T_\lambda^*$  we tacitly assume  $\lambda > 1$ .) For any property  $A$  of rooted trees we let  $\Pr[A]$ ,  $\Pr^*[A]$  denote the probability (as a function of  $\lambda$ ) of  $A$  in  $T_\lambda$ ,  $T_\lambda^*$  respectively.

The *first order logic* for rooted trees consists of equality ( $x = y$ ), parent ( $\pi(x, y)$ , meaning  $x$  is the parent of  $y$ ), the constant symbol  $R$  (the root), the usual Boolean connectives and existential and universal quantification over vertices. A *first order property* is a property that can be written with a sentence  $A$  in this language. The quantifier depth of any first order sentence is the number of nested quantifiers involved in expressing the sentence. We illustrate with a few examples what a typical first order sentence looks like.

**Example 1.1.** *Consider the property that there exists a node in the tree that has precisely two children. This can be expressed in first order language as follows:*

$$\exists u [\exists v_1 [\exists v_2 [\pi(u, v_1) \wedge \pi(u, v_2) \wedge [\forall v \{ \pi(u, v) \implies \{(v = v_1) \vee (v = v_2)\} \}]]]]].$$

*In this particular example, the quantifier depth is 4.*

**Example 1.2.** *Consider the property that the root of the tree has precisely one child and precisely one grandchild. Observe that the root of the tree being a designated symbol, this property is written in first order language as follows:*

$$\exists u [\exists v [\pi(R, u) \wedge \pi(u, v) \wedge [\forall u' \{ \pi(R, u') \implies (u' = u) \}] \wedge [\forall v' \{ \pi(u, v') \implies (v' = v) \}]]].$$

*The quantifier depth is 3.*

We refer the reader to Spencer (2001) for further discussion on first order logic.

Our main results (Theorem 4.7 and Corollary 4.8) will be a characterization of the possible  $\Pr^*[A]$ , as functions of  $\lambda$ , where  $A$  is a first order property. However, it can be shown that the property of  $T$  being infinite is not first order.

---

2010 *Mathematics Subject Classification.* Primary05C20; Secondary60F20.

*Key words and phrases.* Galton-Watson trees, almost sure theory, first order logic.

**Notations 1.3.** Let  $v \in T$ ,  $T$  a rooted tree.  $T(v)$  denotes the subtree of  $T$  that is rooted at  $v$ .  $w$  is an  $i$ -descendant of  $v$  if there is a sequence  $v = x_0, x_1, \dots, x_i = w$  so that  $x_j$  is the parent of  $x_{j+1}$  for  $0 \leq j < i$ . (We say  $v$  is a 0-descendant of itself.) (In the Ulam-Harris notation for trees, this can be expressed as  $w = (x_0, x_1, \dots, x_i)$  where  $x_0 = v$  and  $x_i = w$ .)  $w$  is a  $(\leq i)$ -descendant of  $v$  if it is a  $j$ -descendant for some  $0 \leq j \leq i$ . (E.g., 3-descendants are great-grandchildren.) We define  $d(T)$  to be the depth of the tree, which may be infinite. For  $n \geq 1$ ,  $T|_n$  denotes the first  $n$  generations of  $T$ , along with the root. That is, if  $d(T) > n$ , then we sever all nodes after the  $n$ -th generation (where root is the 0-th generation) and call the truncated tree  $T|_n$ . If, of course,  $d(T) \leq n$ , then  $T|_n = T$ . Let  $T_0$  be a finite tree. We say  $T$  contains  $T_0$  as a subtree if for some  $v \in T$ ,  $T(v) \cong T_0$ . We note that this is a first order property. Letting  $T_0$  have  $s$  nodes, the first order sentence is that there exist distinct  $v_1, \dots, v_s$  having all the desired parent relations and with  $v_1, \dots, v_s$  having no additional children.

We use a *fictitious continuation* to analyze  $T_\lambda$ . Let  $X_1, X_2, \dots$  be a countable sequence of mutually independent and identically distributed  $\text{Poisson}(\lambda)$  random variables. Let  $X_i$  be the number of children of the  $i$ -th node, when the tree is explored using Breadth First Search. (The root is considered the first node so that  $X_1$  is its number of children.) If and when the tree terminates (this occurs when  $\sum_{i=1}^n X_i = n - 1$  for the first time) the remaining (fictitious)  $X_j$  are not used.

**Theorem 1.4.** Fix an arbitrary finite tree  $T_0$ . Consider the following statement  $A$ :

$$A := \{\text{either } T \text{ contains } T_0 \text{ as a subtree or } T \text{ is finite}\}. \quad (1.3)$$

Then  $\Pr[A] = 1$ .

This is one of the main results of this paper. Note that, in particular, Theorem 1.4 immediately implies that for any arbitrary but fixed finite  $T_0$ ,

$$\Pr[\exists v : T(v) \cong T_0] = 1. \quad (1.4)$$

This gives us a good structural description of the infinite random Galton-Watson tree, in the sense that every local neighbourhood is almost surely present somewhere inside the tree.

**1.1. Rapidly Determined Properties.** We say (employing a useful notion of Donald Knuth) that an event is *quite surely determined* in a certain parameter  $s$  if the probability of the complement of that event is exponentially small in  $s$ .

**Definition 1.5.** Consider the fictitious continuation process  $T_\lambda$ . We say that an event  $B$  is *rapidly determined* if *quite surely*  $B$  is *tautologically determined* by  $X_1, X_2, \dots, X_s$ . Here, *tautologically determined* means that for every point  $\omega$  in the sample space, the realization  $(X_1(\omega), X_2(\omega), \dots, X_s(\omega))$  completely determines whether the event  $B$  occurs or not. This means that for every sufficiently large  $s \in \mathbb{N}$ ,

$$\Pr[B \text{ is not determined by } X_1, X_2, \dots, X_s] \leq e^{-\beta s} \quad (1.5)$$

where  $\beta > 0$  is independent of  $s$ .

**Theorem 1.6.** The event  $A$  described in (1.3) is a rapidly determined property.

We shall now prove Theorem 1.4 subject to Theorem 1.6. Fix an arbitrary finite  $T_0$ . Assume Theorem 1.4 is false so that  $\Pr[A] < 1$ , where  $A$  is as in (1.3). For each  $s \in \mathbb{N}$ , with probability at least  $1 - \Pr[A]$  the values  $X_1, \dots, X_s$  do not terminate the tree, nor do they force a copy of  $T_0$ . Then  $A$  would not be tautologically determined. So  $A$  would not be *rapidly determined* and Theorem 1.6 would be false. Taking the contrapositive, Theorem 1.6 implies Theorem 1.4. We prove Theorem 1.6 in §2.1.

*Remark 1.7.* The conclusion of Theorem 1.4 is really that, fixing any finite tree  $T_0$ ,  $T_\lambda^*$  contains  $T_0$  as a subtree with probability one. We can say a bit more. Let  $T_0$  have root  $v$ . For  $L \geq 1$  define  $T_0[L]$  by adding  $L$  new points  $v_0, \dots, v_{L-1}$  and making  $v_i$  a child of  $v_{i-1}$ ,  $1 \leq i \leq L-1$  and  $v$  a child of  $v_{L-1}$ .  $T_\lambda^*$  contains  $T_0[L]$  with probability one. But then it contains a  $T_0$  where the root of  $T_0$  is at least distance  $L$  from the root of  $T$ . We thus deduce that for any finite  $T_0$  and any  $L$  there will, with probability one in  $T_\lambda^*$ , be a  $v$  at distance at least  $L$  from the root such that  $T(v) \cong T_0$ .

**1.2. Ehrenfeucht Games.** We use a very standard and well-known tool to analyze first order properties on rooted trees, namely the Ehrenfeucht games. The Ehrenfeucht games are what bridges the gap between mathematical logic and a complete structural description of logical statements on graphs. Fix a positive integer  $k$ . The standard  $k$ -move Ehrenfeucht game used to analyze first order properties partitions the space of all rooted trees into finitely many equivalence classes. Any two trees belonging to the same equivalence class if and only if they have the same truth value for every first order property of quantifier depth  $\leq k$ . That is, given a first order sentence  $A$  of quantifier depth at most  $k$ , if  $A$  holds true for one of the trees in an equivalence class, then it holds true for all others in that class as well. This notion is made more precise in the following exposition.

We begin with describing the *standard* game, and later move on to a more specialized variant of the game that is suited to our analysis. Fix  $k \geq 1$  and two trees  $T_1$  rooted at  $R_1$  and  $T_2$  rooted at  $R_2$  (these are known to both players). The Ehrenfeucht game  $EHR[T_1, T_2; k]$  is a  $k$ -round game between two players, the Spoiler and the Duplicator. In each round Spoiler picks a vertex from *either*  $T_1$  or  $T_2$  and then Duplicator picks a vertex from the other tree. Letting  $x_1, \dots, x_k; y_1, \dots, y_k$  be the vertices selected (in that order) from  $T_1, T_2$  respectively, Duplicator wins if *all* of the following hold:

- (i)  $x_i = R_1$  iff  $y_i = R_2$ ;
- (ii)  $\pi(x_i, x_j)$  iff  $\pi(y_i, y_j)$ , i.e.  $x_i$  is the parent of  $x_j$  if and only if  $y_i$  is the parent of  $y_j$ ;
- (iii)  $\pi(R_1, x_i)$  iff  $\pi(R_2, y_i)$ , i.e., if  $x_i$  is a child of the root  $R_1$ , then  $y_i$  is a child of  $R_2$ , and vice versa;
- (iv)  $x_i = x_j$  iff  $y_i = y_j$ .

We write  $T_1 \equiv_k T_2$  if and only if Duplicator wins  $EHR[T_1, T_2; k]$ . This equivalence relation partitions all rooted trees into finitely many equivalence classes. It can be shown that two rooted trees  $T_1, T_2$  (with roots  $R_1, R_2$ ) have the same  $k$ -Ehrenfeucht value iff they satisfy precisely the same first order properties of quantifier depth at most  $k$ .

We shall now describe the promised modified version of the game. Let  $T$  be a rooted tree,  $v \in T$ , and  $r > 0$ . Let  $T^-$  be the (undirected) tree on the same vertex set with  $x, y$  adjacent iff one of them is the parent of the other. Let  $B_T(v; r)$  denote the ball of radius  $r$  around  $v$ . That is,

$$B_T(v; r) = \{u \in T : d(u, v) < r \text{ in } T^-\} \quad (1.6)$$

Here  $d(\cdot, \cdot)$  gives the usual graph distance. (For example, cousins are at distance four.)

Let  $k$  (the number of rounds) and  $M$  (an upper bound on the maximal distance) be fixed. Let  $T_1, T_2$  be trees with *designated nodes*  $v_1 \in T_1, v_2 \in T_2$ . Set

$$B_i = B_{T_i}(v_i; \lfloor M/2 \rfloor), \quad i = 1, 2.$$

The  $k$ -move  $M$ -distance preserving Ehrenfeucht game, denoted by  $EHR_M[B_1, B_2; k]$ , is played on these balls. We add a round zero in which the moves  $v_1, v_2$  must be played. (Essentially these are designated vertices.) As before, each round (1 through  $k$ ) Spoiler picks a vertex from either  $T_1$  or  $T_2$  and then Duplicator picks a vertex from the other tree. Letting  $x_0, \dots, x_k; y_0, \dots, y_k$  be the vertices selected from  $T_1, T_2$  respectively, Duplicator wins if

- For  $0 \leq i, j \leq k$ ,  $d(x_i, x_j) = d(y_i, y_j)$ . Equivalently, for all  $1 \leq s \leq M$  and all  $0 \leq i, j \leq k$ ,  $d(x_i, x_j) = s$  if and only if  $d(y_i, y_j) = s$ .
- For  $0 \leq i, j \leq k$ ,  $\pi(x_i, x_j)$  iff  $\pi(y_i, y_j)$ .
- For  $0 \leq i, j \leq k$ ,  $x_i = x_j$  iff  $y_i = y_j$ .

Two balls  $B_1, B_2$  (as described above) are said to have *the same*  $(M; k)$ -Ehrenfeucht value if Duplicator wins  $EHR_M[B_1, B_2; k]$ . We denote this by

$$B_1 \equiv_{M; k} B_2 \quad (1.7)$$

This being an equivalence relation, the space of all such balls with designated centers, is partitioned into  $(M; k)$ -equivalence classes. We let  $\Sigma_{M; k}$  denote the set of all  $(M; k)$ -equivalence classes.

We create a first order language consisting of  $=, \pi(x, y)$  and  $d(x, y) = s$  for  $1 \leq s \leq M$  (note that  $s$  is *not* a variable here). There are only finitely many binary predicates (relations involving two variables). (In general adding the distance function would add an unbounded number of binary predicates. In our case, however, the diameter is bounded by  $M$  and so we are only adding the  $M$  predicates  $d(x, y) = s, 1 \leq s \leq M$ .)

Hence the number of equivalence classes corresponding to this game will also be finite. That is,  $\Sigma_{M;k}$  is a finite set.

**1.3. Universal Trees.** A *universal tree*, as defined below, shall have the property that once  $T$  contains it, all first order statements up to quantifier depth  $k$  depend only on the local neighborhood of the root.

**Definition 1.8.** Fix a positive integer  $k$ . Let

$$M_0 = 2 \cdot 3^{k+1}. \quad (1.8)$$

A finite tree  $T_0$  will be called *universal* if the following phenomenon happens: Take any two trees  $T_1, T_2$  with roots  $R_1, R_2$  such that:

- (i) the  $3^{k+1}$  neighbourhoods around the root have the same  $(M_0; k)$  value, i.e.

$$B_{T_1}(R_1; 3^{k+1}) \equiv_{M_0; k} B_{T_2}(R_2; 3^{k+1}). \quad (1.9)$$

- (ii) For some  $u_1 \in T_1, u_2 \in T_2$  such that

$$d(R_1, u_1) > 3^{k+2}, \quad d(R_2, u_2) > 3^{k+2}, \quad (1.10)$$

we have

$$T_1(u_1) \cong T_2(u_2) \cong T_0. \quad (1.11)$$

Then  $T_1 \equiv_k T_2$ . Equivalently, Duplicator wins the *standard*  $k$ -move Ehrenfeucht game played on  $T_1, T_2$ .

*Remark 1.9.* Technically, we should call such a  $T_0$  as described in Definition 1.8 *k-universal*. However, in the sequel, we simply refer to this as *universal* for the convenience of notation, and since the dependence on  $k$  will be clear in each context.

We prove in Theorem 3.3 that such a *universal tree* indeed exists, by imposing sufficient structural conditions on it.

*Remark 1.10.* Fix a certain *universal tree*  $UNIV_k$ , given  $k \in \mathbb{N}$ . Using theorem 1.4, we conclude that  $T_\lambda^*$  will almost surely contain  $UNIV_k$ . From Remark 1.7, we say further that there will almost surely exist a node  $v$  at distance  $> 3^{k+2}$  from the root such that

$$T(v) \cong UNIV_k.$$

From the definition of *universal trees*, then the *standard* Ehrenfeucht value of  $T_\lambda^*$  will be determined by the  $(M_0; k)$ -Ehrenfeucht value of  $B_{T_\lambda^*}(R; 3^{k+1})$ , the  $3^{k+1}$ -neighbourhood of the root  $R$ , where  $M_0$  is as in (1.8).

**1.4. An Almost Sure Theory.** Let  $\mathcal{B}_i, 1 \leq i \leq N$  for some positive integer  $N$ , denote the finitely many  $(M_0; k)$ -equivalence classes. Note that these are defined on balls of radius  $3^{k+1}$  centered at a designated vertex which is a node in some tree. Then for every realization  $T$  of  $T_\lambda^*$ ,

$$B_T(R; 3^{k+1}) \in \mathcal{B}_i \quad \text{for precisely one } i, 1 \leq i \leq N. \quad (1.12)$$

Almost surely for two realizations  $T_1, T_2$  of  $T_\lambda^*$  which have the same local neighbourhoods of the roots, i.e.

$$B_{T_1}(R_1; 3^{k+1}) \in \mathcal{B}_i, \quad B_{T_2}(R_2; 3^{k+1}) \in \mathcal{B}_i \quad \text{for the same } i,$$

we have  $T_1 \equiv_k T_2$ . As the  $\mathcal{B}_i$  are equivalence classes over the space of rooted trees they may be considered properties of rooted trees and so have probabilities  $\Pr^*[\mathcal{B}_i]$  in  $T_\lambda^*$ . As they finitely partition the space of all rooted trees

$$\sum_{i=1}^N \Pr^*[\mathcal{B}_i] = 1. \quad (1.13)$$

Let  $\mathcal{AS}$  denote the almost sure theory for  $T_\lambda^*$ . That is,  $\mathcal{AS}$  consists of all first order sentences  $B$  such that  $\Pr^*[B] = 1$ . We now give an axiomatization of  $\mathcal{AS}$ . Let  $\mathcal{T}$  be defined by the following schema:

$$A[T_0] := \{\exists v : T(v) \cong T_0\}, \quad \text{for all } T_0 \text{ finite trees.} \quad (1.14)$$

**Theorem 1.11.** *Under the probability  $\Pr^*$ ,*

$$\mathcal{T} = \mathcal{AS} \quad (1.15)$$

*That is, the first order statements  $B$  with  $\Pr^*[B] = 1$  are precisely those statements derivable from the axiom schema  $\mathcal{T}$ .*

As  $\mathcal{T}$  does not depend on  $\lambda$  we also have:

**Corollary 1.12.** *The almost sure theory  $\mathcal{AS}$  is the same for all  $\lambda > 1$ .*

That  $\mathcal{T} \subseteq \mathcal{AS}$  is already clear from Theorem 1.4. To show the reverse inclusion, consider for every  $1 \leq i \leq N$ ,  $\mathcal{T} + \mathcal{B}_i$ . In this theory every finite  $T_0$  is contained as a subtree and the  $3^{k+1}$ -neighbourhood of the root belongs to the equivalence class  $\mathcal{B}_i$ . As discussed above in Remark 1.10, this set of information completely determines the *standard* Ehrenfeucht value of the infinite tree. That is, for any first order sentence  $A$  of quantifier depth  $k$

$$\text{either } \mathcal{T} + \mathcal{B}_i \models A \quad \text{or } \mathcal{T} + \mathcal{B}_i \models \neg A. \quad (1.16)$$

The standard notation  $T \models A$  for a tree  $T$  and a property  $A$  means that the property  $A$  holds true for tree  $T$ . Set

$$J_A = \{1 \leq i \leq N : \mathcal{T} + \mathcal{B}_i \models A\}. \quad (1.17)$$

Under  $T_\lambda^*$ ,  $A$  holds *if and only if*  $\mathcal{B}_i$  holds for some  $i \in J_A$ . Thus we can express

$$\Pr^*[A] = \sum_{i \in J_A} \Pr^*[\mathcal{B}_i]. \quad (1.18)$$

In Section 4 we shall use this to express all  $\Pr^*[A]$  in reasonably succinct form.

Now suppose, under  $T_\lambda^*$ , that  $\Pr^*[A] = 1$ . As the  $\mathcal{B}_i$  partition the neighbourhoods around the roots of trees, this implies that  $J_A = \{1, 2, \dots, N\}$ . That is,  $\mathcal{T} + \mathcal{B}_i \models A$  for all  $1 \leq i \leq N$  and  $\bigvee_{i=1}^N \mathcal{B}_i$  is a tautology. Hence  $A$  is derivable from  $\mathcal{T}$ . Thus  $\mathcal{AS} \subseteq \mathcal{T}$ .

In Section 4 below, we turn to the computation of the possible  $\Pr^*[A]$ . As seen above, in the space of  $T_\lambda^*$ , the neighbourhoods around the root of sufficiently large radius are instrumental in determining the *standard* Ehrenfeucht value of the tree. It only makes sense, therefore, to compute the probabilities of having specific neighbourhoods around the root conditioned on the tree being infinite. We shall do this in a recursive fashion, using induction on the number of generations below the root that we are considering.

## 2. CONTAINING ALL FINITE TREES

**2.1. A Rapidly Determined Property.** We prove here Theorem 1.6. We fix an arbitrary finite tree  $T_0$  with depth  $d(T_0) = d_0$ , following the notation given in 1.3. We alter the *fictitious continuation* process  $T_\lambda$  described previously. If for some finite, first  $n \in \mathbb{N}$ , we have  $\sum_{i=1}^n X_i = n - 1$ , then the actual tree has vertices  $1, \dots, n$ . If this phenomenon does not happen for any finite  $n$ , then we have one infinite tree described by our fictitious continuation. If the tree does abort after  $n$  vertices, we begin a new tree with vertex  $n + 1$  as the root, and generate it from  $X_{n+1}, X_{n+2}, \dots$ . Again, if this tree terminates at some  $n_1$  we begin a new tree with vertex  $n_1 + 1$ . Continuing, we generate an infinite forest, with vertices the positive integers. We call this the *forest* process and label it  $T_\lambda^{for}$ .

Then we define, for every  $s \in \mathbb{N}$ , the event (in  $T_\lambda$ )

$$good(s) = \{A \text{ is completely determined by } X_1, \dots, X_s\}, \quad (2.1)$$

where  $A$  is as in (1.3). Set  $bad(s) = good(s)^c$ . For every node  $i \in \mathbb{N}$ , define in  $T_\lambda^{for}$

$$I_i = \mathbf{1}_{T(i) \cong T_0}. \quad (2.2)$$

That is,  $I_i$  is the indicator function of the event that *in the random forest*  $T(i) \cong T_0$ . Set

$$Y = \sum_{i=1}^{\lfloor \epsilon^{d_0} s \rfloor} I_i, \quad (2.3)$$

where, with foresight, we require

$$0 < \epsilon < \frac{1}{\lambda + 1}. \quad (2.4)$$

(Our  $\epsilon$  is chosen sufficiently small so that quite surely in  $s$ , in  $T_\lambda^{for}$ , all of the  $(\leq d_0)$ -descendants  $j$  of all  $i \leq s\epsilon$  have  $j \leq s$ .) We create a martingale, setting, for  $1 \leq i \leq s$ ,

$$Y_i = E[Y | X_1, X_2, \dots, X_i], \quad Y_0 = E[Y]. \quad (2.5)$$

In  $T_\lambda^{for}$ , for  $x \in \mathbb{R}^+$ ,  $i \in \mathbb{N}$ , set

$$\mathcal{S}_i(x) = \{\text{indices of all } i\text{-descendants of nodes } 1, 2, \dots, \lfloor x \rfloor\} \quad (2.6)$$

with  $\mathcal{S}_0(x) = \{1, 2, \dots, \lfloor x \rfloor\}$ , where an  $i$ -descendant is as described in Notations 1.3. Define, for  $i \in \mathbb{N}$ ,

$$g_i(x) = \text{highest index recorded in } \bigcup_{j=0}^i \mathcal{S}_j(x). \quad (2.7)$$

**Lemma 2.1.** *For any  $x \in \mathbb{R}^+$ ,  $d \in \mathbb{N}$ ,*

$$g_d(x) = g_1^d(x). \quad (2.8)$$

*Here  $g_1^d$  denotes the  $d$ -times composition of  $g_1$ .*

*Proof.* We prove this using induction on  $d$ . For  $d = 1$  this is true by definition of  $g_1$ . For  $d = 2$ , the highest possible index of all the children and grandchildren of  $1, 2, \dots, \lfloor x \rfloor$  is equal to the highest index of the children of the nodes  $1, 2, \dots, g_1(\lfloor x \rfloor) = g_1(x)$ , which is  $g_1(g_1(x))$ . Now suppose we have proved the claim for some  $d \in \mathbb{N}$ ,  $d \geq 2$ . Once again, a similar argument comes into play. The highest index among all the  $(d+1)$ -descendants of nodes  $1, 2, \dots, \lfloor x \rfloor$ , is also equal to the highest index among all the  $d$ -descendants of the nodes  $1, 2, \dots, g_1(x)$ , which by induction hypothesis will be  $g_1^d(g_1(x)) = g_1^{d+1}(x)$ .  $\square$

When  $g_{d_0}(\lfloor \epsilon^{d_0} s \rfloor) \leq s$ , the descendants  $j$  of  $1, \dots, \lfloor \epsilon^{d_0} s \rfloor$  down to generation  $d_0$  will all have  $j \leq s$ . Thus  $Y$  will be completely determined by  $X_1, \dots, X_s$ . That is,

$$g_1^{d_0}(\lfloor \epsilon^{d_0} s \rfloor) \leq s \quad \Rightarrow \quad Y_s = Y. \quad (2.9)$$

A few observations about the function  $g_1(\cdot)$  are important. First,

$$g_1(x) \geq \lfloor x \rfloor \quad \text{for all } x \in \mathbb{R}^+. \quad (2.10)$$

In  $T_\lambda^{for}$  every time the tree terminates, we start a new tree, and that uses up one extra index for the root of the new tree. But while counting the nodes  $1, 2, \dots, \lfloor x \rfloor$ , for any  $x \in \mathbb{R}^+$ , at most  $\lfloor x \rfloor$  many new trees need be started. Therefore

$$g_1(x) \leq \lfloor x \rfloor + \sum_{i=1}^{\lfloor x \rfloor} X_i. \quad (2.11)$$

Further, by the definition of  $g_1(\cdot)$ , it is clear that it is monotonically increasing.

We shall use the inequality in (2.11) to show that, for  $\epsilon$  as chosen in (2.4), quite surely in  $s$ , we have  $Y_s = Y$ , i.e.  $Y$  is tautologically determined by  $X_1, \dots, X_s$  with exponentially small failure probability in  $s$ . This involves showing that for  $i$  this small, i.e.  $1 \leq i \leq \lfloor \epsilon^{d_0} s \rfloor$ ,  $T(i)$  is quite surely determined by  $X_1, \dots, X_s$ .

We employ Chernoff bounds. For  $x \in \mathbb{R}^+$  and any  $\alpha > 0$ ,

$$\begin{aligned} \Pr[g_1(\epsilon x) > x] &= \Pr[e^{\alpha g_1(\epsilon x)} > e^{\alpha x}] \\ &\leq E[e^{\alpha g_1(\epsilon x)}] e^{-\alpha x} \\ &\leq E[e^{\alpha(\epsilon x + \sum_{i=1}^{\lfloor \epsilon x \rfloor} X_i)}] e^{-\alpha x} \end{aligned}$$

$$\begin{aligned}
&= e^{\alpha \epsilon x} \prod_{i=1}^{\lfloor \epsilon x \rfloor} E[e^{\alpha X_i}] e^{-\alpha x} \\
&= e^{-(1-\epsilon)\alpha x} \{\exp[\lambda(e^\alpha - 1)]\}^{\lfloor \epsilon x \rfloor} \\
&\leq e^{-(1-\epsilon)\alpha x} \{\exp[\lambda(e^\alpha - 1)]\}^{\epsilon x} \\
&= \exp\{-(1-\epsilon)\alpha - \lambda(e^\alpha - 1)\epsilon\}x.
\end{aligned} \tag{2.12}$$

It can be checked that for any  $\alpha \in (0, \log(\frac{1-\epsilon}{\lambda\epsilon}))$ , the exponent in (2.12) is negative, i.e.  $-[(1-\epsilon)\alpha - \lambda(e^\alpha - 1)\epsilon] < 0$ . We set

$$\eta = [(1-\epsilon)\alpha - \lambda(e^\alpha - 1)\epsilon] \tag{2.13}$$

Observe that  $\eta$  is positive. Now we have the upper bound:

$$\Pr[g_1(\epsilon x) > x] \leq e^{-\eta x}. \tag{2.14}$$

We make the following claim:

**Lemma 2.2.** *For any  $d, s \in \mathbb{N}$ ,*

$$\Pr[g_1^d(\epsilon^d s) > s] \leq \sum_{i=0}^{d-1} e^{-\epsilon^i \eta s}. \tag{2.15}$$

*Proof.* We prove this using induction on  $d$ . We have already seen that this holds for  $d = 1$ . This initiates the induction hypothesis. Suppose it holds for some  $d \in \mathbb{N}$ . Then

$$\begin{aligned}
\Pr[g_1^{d+1}(\epsilon^{d+1} s) > s] &= \Pr[g_1^{d+1}(\epsilon^{d+1} s) > s, g_1(\epsilon^{d+1} s) > \epsilon^d s] + \Pr[g_1^{d+1}(\epsilon^{d+1} s) > s, g_1(\epsilon^{d+1} s) \leq \epsilon^d s] \\
&\leq \Pr[g_1(\epsilon^{d+1} s) > \epsilon^d s] + \Pr[g_1^d(\epsilon^d s) > s] \\
&\leq e^{-\eta \cdot \epsilon^d s} + \sum_{i=0}^{d-1} e^{-\epsilon^i \eta s}, \quad \text{by induction hypothesis and (2.14);} \\
&= \sum_{i=0}^d e^{-\epsilon^i \eta s}.
\end{aligned}$$

This completes the proof.  $\square$

From Lemma 2.2, we conclude that

$$\Pr[g_1^{d_0}(\lfloor \epsilon^{d_0} s \rfloor) > s] \leq \sum_{i=0}^{d_0-1} e^{-\epsilon^i \eta s}, \tag{2.16}$$

From (2.9), this means

$$\Pr[Y_s = Y] \geq 1 - \sum_{i=0}^{d_0-1} e^{-\epsilon^i \eta s}. \tag{2.17}$$

As promised earlier, we therefore have that, quite surely,  $Y_s = Y$ . In the following definition, we describe the event  $Y_s = Y$  as *globalgood*( $s$ ), emphasizing the dependence on the parameter  $s$ . What we can conclude from the above computation is that *globalgood*( $s$ ) fails to happen with only exponentially small failure probability in  $s$ .

**Definition 2.3.** *globalgood*( $s$ ) is the event  $Y_s = Y$ . *globalbad*( $s$ ) is the complement of *globalgood*( $s$ ).

We now claim that the martingale  $\{Y_i : 0 \leq i \leq s\}$  satisfies a Lipschitz Condition.

**Lemma 2.4.** *There exists constant  $C > 0$  such that for  $1 \leq i \leq s$ ,*

$$|Y_i - Y_{i-1}| \leq C. \tag{2.18}$$



*Proof.* For  $1 \leq i \leq \lfloor \epsilon^{d_0} s \rfloor$ , fix a sequence  $\vec{x} = (x_1, \dots, x_{i-1}) \in (\mathbb{N} \cup \{0\})^{i-1}$ , and then consider

$$y_i = E[Y|X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_i] = \sum_{1 \leq j \leq \lfloor \epsilon^{d_0} s \rfloor} E[I_j|X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_i]$$

and

$$y_{i-1} = E[Y|X_1 = x_1, \dots, X_{i-1} = x_{i-1}] = \sum_{1 \leq j \leq \lfloor \epsilon^{d_0} s \rfloor} E[I_j|X_1 = x_1, \dots, X_{i-1} = x_{i-1}].$$

$I_j$  will be affected by the extra information about  $X_i$  only if either  $j = i$  or node  $j$  is an ancestor of node  $i$  at distance  $\leq d_0$  from  $i$ . If  $j = i$ , then it will of course affect the conditional expectation because  $X_i$  gives the number of children of  $j$  in that case. When  $j > i$ , this is immediate, because any subtree rooted at  $j$  has no involvement of  $X_i$ . When  $j < i$ , but not an ancestor of  $i$ ,  $i$  is not a part of the subtree  $T(j)$  rooted at  $j$ . Therefore  $X_i$ , the number of children of node  $i$ , does not contribute anything to the probability of the presence of  $T_0$  rooted at  $j$ . When  $j$  is an ancestor of  $i$  but at distance  $> d_0$  from  $i$ ,  $i$  won't be a part of the subtree  $T(j)|_{d_0}$  at all.

When  $j$  is an ancestor of  $i$  and at distance  $d_0$  from  $i$ , then  $i$  is a leaf node of  $T(j)|_{d_0}$  and therefore  $X_i$ , the number of children of  $i$ , will actually play a role, because to ensure that  $T(j) \cong T_0$ , the leaf nodes of  $T(j)|_{d_0}$  must have no children of their own in  $T_\lambda^{for}$ .

That is, we need be concerned with the at most  $d_0$  ancestors of node  $i$ , plus  $i$  itself, and for each of them, the difference in the conditional expectations of  $I_j$  can be at most 1. Denoting by  $\sum^*$  the sum over  $j = i$  and  $j$  an ancestor of  $i$  at distance  $\leq d_0$  from  $i$ , this gives us:

$$\begin{aligned} |y_i - y_{i-1}| &= \left| \sum^* E[I_j|X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_i] - E[I_j|X_1 = x_1, \dots, X_{i-1} = x_{i-1}] \right| \\ &\leq \sum^* |E[I_j|X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_i] - E[I_j|X_1 = x_1, \dots, X_{i-1} = x_{i-1}]| \\ &\leq d_0 + 1. \end{aligned}$$

The final inequality follows from the argument above that  $\sum^*$  involves summing over at most  $d_0 + 1$  many terms, and each summand is at most 1, since each summand is the difference of the expectations of indicator random variables. This proves Lemma 2.4, with  $C = d_0 + 1$ . □

Given Lemma 2.4 we apply Azuma's inequality. Consider the martingale

$$Y'_i = \frac{E[Y] - Y_i}{d_0 + 1}, \quad 0 \leq i \leq s.$$

Set, for a typical node  $v$  in a random Galton-Watson tree  $T$  with *Poisson*( $\lambda$ ) offspring distribution,

$$\Pr[T(v) \cong T_0] = p_0, \tag{2.19}$$

so that  $E[Y] = \lfloor \epsilon^{d_0} s \rfloor p_0$ . Applying Azuma's inequality to  $\{Y'_i, 0 \leq i \leq s\}$ , for any  $\beta > 0$ ,

$$\Pr[Y'_s > \beta \sqrt{s}] < e^{-\beta^2}.$$

We choose

$$\beta = \frac{\epsilon^{d_0} p_0 \sqrt{s}}{2(d_0 + 1)}.$$

This gives

$$\Pr[Y_s < \frac{\epsilon^{d_0} p_0}{2} \cdot s - p_0] < \exp \left\{ -\frac{\epsilon^{2d_0} p_0^2}{4(d_0 + 1)^2} \cdot s \right\}. \tag{2.20}$$

Writing

$$\xi = \frac{\epsilon^{d_0} p_0}{2}, \quad \varphi = \frac{\epsilon^{2(d_0)} p_0^2}{4(d_0 + 1)^2},$$



we can rewrite the above inequality as

$$\Pr[Y_s < \xi s - p_0] < e^{-\varphi s}. \quad (2.21)$$

Putting everything together, we get for all  $s$  large enough:

$$\begin{aligned} \Pr[Y = 0] &= \Pr[Y = 0, Y_s = Y] + \Pr[Y = 0, Y_s \neq Y] \\ &\leq \Pr[Y_s < \xi s - p_0] + \Pr[Y_s \neq Y] \\ &\leq e^{-\varphi s} + \sum_{i=0}^{d_0-2} e^{-\epsilon^i \eta s}; \quad \text{from (2.17) and (2.21);} \end{aligned}$$

which is an upper bound exponentially small in  $s$ . This gives us the proof of Theorem 1.6.

### 3. UNIVERSAL TREES EXIST!

In this section, we shall establish sufficient conditions that guarantee the existence of *universal trees*. Fixing  $k \in \mathbb{N}$ , set  $M_0 = 2 \cdot 3^{k+1}$  as in (1.8). Assume  $T_0$  is a finite tree with root  $R_0$  with the following properties:

- (i) For every  $\sigma \in \Sigma_{M_0; k}$ , there are distinct nodes  $v_{i; \sigma} \in T_0, 1 \leq i \leq k$ , with the following conditions satisfied: for every  $\sigma \in \Sigma_{M_0; k}$  and every  $1 \leq i \leq k$ , we have

$$d(R_0, v_{i; \sigma}) > 3^{k+2}; \quad (3.1)$$

for every  $\sigma_1, \sigma_2 \in \Sigma_{M_0; k}$  and  $1 \leq i_1, i_2 \leq k$ , with  $(\sigma_1, i_1) \neq (\sigma_2, i_2)$ , we have

$$d(v_{i_1; \sigma_1}, v_{i_2; \sigma_2}) > 3^{k+2}; \quad (3.2)$$

and for all  $1 \leq i \leq k, \sigma \in \Sigma_{M_0; k}$ ,

$$B(v_{i; \sigma}; 3^{k+1}) \in \sigma. \quad (3.3)$$

- (ii) For every  $1 \leq i \leq k$ , every choice of  $u_1, \dots, u_{i-1} \in T_0$ , and every choice of  $\sigma \in \Sigma_{M_0; k}$ , there exists a vertex  $u_i \in T_0$  such that

$$d(u_i, u_j) > 3^{k+2}, \text{ for all } 1 \leq j \leq i-1, \quad (3.4)$$

$$d(R_0, u_i) > 3^{k+2}, \quad (3.5)$$

and

$$B(u_i; 3^{k+1}) \in \sigma. \quad (3.6)$$

*Remark 3.1.* Observe that Condition (ii) is stronger than Condition (i) and actually implies the latter. However, for pedagogical clarity, and since (i) gives a nice structural description of the *Christmas tree* that is described in Theorem 3.3, we retain (i). Furthermore, we state (i) before (ii) since, we feel, it is an easier condition to visualize.

**Lemma 3.2.**  $T_0$  with properties described above will be a universal tree.

*Proof.* Recall the definition of universal trees. We start with two trees  $T_1, T_2$  with roots  $R_1, R_2$ , and which satisfy the following conditions:

- (i) The balls  $B(R_1; 3^{k+1}), B(R_2; 3^{k+1})$  satisfy

$$B(R_1; 3^{k+1}) \equiv_{M_0; k} B(R_2; 3^{k+1}). \quad (3.7)$$

- (ii) For some  $u_1 \in T_1, u_2 \in T_2$  such that

$$d(R_1, u_1) > 3^{k+2}, \quad d(R_2, u_2) > 3^{k+2}, \quad (3.8)$$

we have each of  $T_1(u_1)$  and  $T_2(u_2)$  isomorphic to  $T_0$ . If  $\varphi_1 : T_0 \rightarrow T_1(u_1), \varphi_2 : T_0 \rightarrow T_2(u_2)$  are these isomorphisms, then

$$\varphi_1(v_{i; \sigma}) = v_{i; \sigma}^{(1)}, \quad \varphi_2(v_{i; \sigma}) = v_{i; \sigma}^{(2)},$$

for all  $\sigma \in \Sigma_{M_0; k}, 1 \leq i \leq k$ .

Now we give a winning strategy for the Duplicator. We assume that since  $R_1, R_2$  are designated vertices,  $x_0 = R_1, y_0 = R_2$ . Let  $(x_i, y_i)$  be the pair chosen from  $T_1 \times T_2$  in the  $i$ -th move, for  $1 \leq i \leq k$ . Now, we claim the following:

The Duplicator can play the game such that, for each  $0 \leq i \leq k$ ,

- he can maintain

$$B(x_i; 3^{k+1-i}) \equiv_{M_0; k} B(y_i; 3^{k+1-i}),$$

(Our proof only needs

$$B(x_i; 3^{k+1-i}) \equiv_{M_0; k-i} B(y_i; 3^{k+1-i}),$$

but the stronger assumption is a bit more convenient);

- for all  $0 \leq j < i$  such that  $x_j \in B(x_i; 3^{k+1-i})$ , the corresponding  $y_j \in B(y_i; 3^{k+1-i})$ , and vice versa, according to the winning strategy of  $EHR_{M_0}[B(x_i; 3^{k+1-i}), B(y_i; 3^{k+1-i}); k]$ . Again, this is overkill as one need only consider the Ehrenfeucht game of  $k - i$  moves at this point.

We prove this using induction on the number of moves played so far. For  $i = 0$ , we have chosen  $x_0 = R_1, y_0 = R_2$ , and we already have imposed the condition

$$B(R_1; 3^{k+1}) \equiv_{M_0; k} B(R_2; 3^{k+1})$$

in (3.7). So suppose the claim holds for  $0 \leq j \leq i - 1$ . Without loss of generality suppose Spoiler chooses  $x_i \in T_1$ . There are two possibilities:

- (i) *Inside move:*

$$x_i \in \bigcup_{j=0}^{i-1} B(x_j; 2 \cdot 3^{k+1-i}). \quad (3.9)$$

So  $x_i \in B(x_l; 2 \cdot 3^{k+1-i})$  for some  $0 \leq l \leq i - 1$ . By the induction hypothesis,

$$B(x_l; 3^{k+1-l}) \equiv_{M_0; k} B(y_l; 3^{k+1-l}).$$

Duplicator now follows his winning strategy of  $EHR_{M_0}[B(x_l; 3^{k+1-l}), B(y_l; 3^{k+1-l}); k]$  and picks  $y_i \in B(y_l; 3^{k+1-l})$ . This means that,

$$d(x_i, x_l) < 2 \cdot 3^{k+1-i} \Rightarrow B(x_i; 3^{k+1-i}) \subset B(x_l; 3^{k+1-l}),$$

since  $l < i$ . In the same way

$$B(y_i; 3^{k+1-i}) \subset B(y_l; 3^{k+1-l}),$$

and further,

$$B(x_i; 3^{k+1-i}) \equiv_{M_0; k} B(y_i; 3^{k+1-i}).$$

This last relation follows from the fact that  $y_i$  is chosen corresponding to  $x_i$  in the winning strategy of the Duplicator for  $EHR_{M_0}[B(x_l; 3^{k+1-l}), B(y_l; 3^{k+1-l}); k]$ . Since  $M_0$ , as chosen in Equation (1.8), is greater than  $2 \cdot 3^{k+1-i}$ , hence for Duplicator to win  $EHR_{M_0}[B(x_l; 3^{k+1-l}), B(y_l; 3^{k+1-l}); k]$ , he must be able to win the game played within the smaller balls  $B(x_i; 3^{k+1-i})$  and  $B(y_i; 3^{k+1-i})$ .

- (ii) *Outside move:*

$$x_i \notin \bigcup_{j=0}^{i-1} B(x_j; 2 \cdot 3^{k+1-i}). \quad (3.10)$$

Then we consider  $B(x_i; 3^{k+1-i})$  and we know, from (3.4), (3.5) and (3.6), that there exists some  $v \in T_2$  such that

$$d(v, y_l) > 3^{k+2}, \quad \text{for all } 0 \leq l \leq i - 1,$$

and

$$B(v; 3^{k+1}) \equiv_{M_0; k} B(x_i; 3^{k+1}).$$

We choose  $y_i = v$ . Note that then we automatically have

$$B(y_i; 3^{k+1-i}) \cap \left\{ \bigcup_{j=0}^{i-1} B(y_j; 3^{k+1-i}) \right\} = \phi,$$

and

$$B(x_i; 3^{k+1-i}) \equiv_{M_0; k} B(y_i; 3^{k+1-i}).$$

Once again, Duplicator is choosing  $y_i$  so that  $B(y_i; 3^{k+1}) \equiv_{M_0; k} B(x_i; 3^{k+1})$ , i.e. he wins

$$EHR_{M_0} [B(x_i; 3^{k+1}), B(y_i; 3^{k+1}); k].$$

Then he must be able to win the game within the smaller balls  $B(x_i; 3^{k+1-i})$  and  $B(y_i; 3^{k+1-i})$ , since his winning involves being able to preserve mutual distances of pairs of nodes up to  $M_0$ .

This shows that the Duplicator will win  $EHR[T_1, T_2; k]$ , which finishes the proof.  $\square$

**Theorem 3.3.** *For each  $k \in \mathbb{N}$  there is a universal tree  $T$ .*

*Proof.*  $T$  will be a *Christmas tree* which is constructed as follows. For each  $\sigma \in \Sigma_{M_0; k}$  select and fix a specific ball  $B(v; 3^{k+1}) \in \sigma$ . For each such  $\sigma$  and each  $1 \leq i \leq k$  create disjoint copies  $T_{i, \sigma} = B(v_{i, \sigma}; 3^{k+1})$  such that  $B(v_{i, \sigma}; 3^{k+1}) \cong B(v; 3^{k+1})$ , with the isomorphism mapping  $v_{i, \sigma}$  to  $v$ . These  $B(v_{i, \sigma}; 3^{k+1})$  are the *balls* decorating the Christmas tree. Let  $w_{i, \sigma}$  be the *top* vertex of  $B(v_{i, \sigma}; 3^{k+1})$ . That is, it is that unique node in the ball with no ancestor in the ball. It can be seen that this node is actually the ancestor of  $v_{i, \sigma}$  which is at distance  $3^{k+1}$  away from  $v_{i, \sigma}$ , or in other words,  $v_{i, \sigma}$  is a  $3^{k+1}$ -descendant of this node. Let  $R$  be the root of  $T$ . Draw disjoint paths of length  $3^{k+4}$  from  $R$  to each  $w_{i, \sigma}$ . These will be like the *strings* attaching the balls to the Christmas tree.

We now explain why this  $T$  satisfies Conditions (i) and (ii). Once again, for pedagogical clarity, we first show a detailed reasoning why  $T$  satisfies (i), although technically, it suffices to verify only (ii). First, observe that the  $v_{i, \sigma}$  we have defined in the previous paragraph, for  $1 \leq i \leq k$  and  $\sigma \in \Sigma_{M_0; k}$ , immediately satisfy (3.1) and (3.2), since

$$d(R, v_{i, \sigma}) = d(R, w_{i, \sigma}) + d(v_{i, \sigma}, w_{i, \sigma}) = 3^{k+4} + 3^{k+1} > 3^{k+2},$$

for every  $\sigma_1, \sigma_2 \in \Sigma_{M_0; k}$ ,  $1 \leq i_1, i_2 \leq k$  with  $(\sigma_1, i_1) \neq (\sigma_2, i_2)$ , we indeed have

$$d(v_{i_1, \sigma_1}, v_{i_2, \sigma_2}) = d(v_{i_1, \sigma_1}, R) + d(R, v_{i_2, \sigma_2}) > 2 \cdot 3^{k+4} > 3^{k+2}.$$

To see that (3.3) holds, note that by our construction,

$$B(v_{i, \sigma}; 3^{k+1}) \cong B(v; 3^{k+1}) \in \sigma,$$

with  $v_{i, \sigma}$  mapped to  $v$ , for all  $1 \leq i \leq k$ , and for all  $\sigma \in \Sigma_{M_0; k}$ .

Finally, we verify that (ii) holds. Consider any  $1 \leq j \leq k$ . Suppose we have selected any  $j-1$  vertices  $u_1, \dots, u_{j-1}$  from  $T$ . For any  $\sigma \in \Sigma_{M_0; k}$  and  $1 \leq i \leq k$ , we consider the *branch* of the tree consisting of the ball  $B(v_{i, \sigma}; 3^{k+1})$  and the string joining  $R$  to  $w_{i, \sigma}$ , and we call that branch *free* if no  $u_l$ ,  $1 \leq l \leq j-1$  is picked from that branch. Since there are  $k$  copies of balls for each  $\sigma$ , and  $j \leq k$ , hence we shall always have at least one *free* branch from each  $\sigma \in \Sigma_{M_0; k}$ . So we simply choose  $u_j = v_{i, \sigma}$  for some  $i$  such that the corresponding branch is free.

Since no  $u_l$ ,  $1 \leq l \leq j-1$ , belongs to that branch, each of them must be at least as far away from  $u_j$  as the root is from  $v_{i, \sigma}$ . That is, we will have

$$d(u_j, u_l) > 3^{k+4} + 3^{k+1}; \quad d(u_j, R) = 3^{k+4} + 3^{k+1}.$$

And of course, by our choice, we would have  $B(u_j; 3^{k+1}) \in \sigma$ .  $\square$

## 4. PROBABILITIES CONDITIONED ON INFINITENESS OF THE TREE

As before, with  $R$  the root,  $B_T(R; i)$  denotes the neighbourhood of  $R$  with radius  $i$ , i.e.

$$B_T(R; i) = \{u \in T : d(u, R) < i\}.$$

We define

$$\overline{B_T(R; i)} = \{u \in T : d(u, R) \leq i\}.$$

So,  $\overline{B_T(R; i)}$  captures up to the  $i$ -th generation of the tree,  $R$  being the 0-th generation. For each  $i \in \mathbb{N}$  we give a set of equivalence classes  $\Gamma_i$  which will be relatively easy to handle and which we show in Theorem 4.2 is a refinement of  $\Sigma_{i:k}$ . We set

$$C = \{0, 1, \dots, k-1, \omega\}. \quad (4.1)$$

Here  $\omega$  is a special symbol with the meaning “at least  $k$ .” That is, to say that there are  $\omega$  copies of something is to say that there are at least  $k$  copies. We set

$$\Gamma_1 = C = \{0, 1, \dots, k-1, \omega\}. \quad (4.2)$$

A  $\overline{B_T(R; 1)}$  is of type  $i \in \Gamma_1$  if the root has  $i$  children. Since the game has  $k$  rounds, if the roots has  $x, y$  children in the two trees with both  $x, y \geq k$  then Duplicator wins the modified game. Inductively we now set

$$\Gamma_{i+1} = \{g : \Gamma_i \rightarrow C\}. \quad (4.3)$$

Each child  $v$  of the root generates a tree to generation  $i$ . This tree belongs to an equivalence class  $\sigma \in \Gamma_i$ . A  $\overline{B_T(R; i+1)}$  has state  $g \in \Gamma_{i+1}$  if for all  $\sigma \in \Gamma_i$  the root has  $g(\sigma)$  children  $v$  whose subtree  $T(v)$  upto generation  $i$  belongs to equivalence class  $\sigma$ , i.e.  $T(v)|_i \in \sigma$ .

**Example 4.1.** Consider  $k = 4, i = 2$ . Then a typical example of  $\overline{B_T(R; i)}$  will be: the root has two children with no children, at least four children with one child, three children with two children, no children with three children, and one child with at least four children. Thus  $g(0) = 2, g(1) = \omega, g(2) = 3, g(3) = 0, g(\omega) = 1$ .

**Theorem 4.2.**  $\Gamma_i$  is a refinement on  $\Sigma_{i:k}$ .

*Proof.* Let  $\overline{B_{T_1}(R_1; i)}, \overline{B_{T_2}(R_2; i)}$  lie in the same  $\Gamma_i$  equivalence class. It suffices to show that Duplicator wins the  $k$ -move modified Ehrenfeucht game on these balls. We show this using induction on  $i$ .

The case  $i = 1$  is immediate. Suppose it holds good for all  $i' \leq i-1$ . In the Ehrenfeucht game let Spoiler select  $w_1 \in T_1$ . Let  $v_1$  be the child of the root such that  $w_1$  belongs to the tree generated by  $v_1$  up to depth  $i-1$ , i.e.  $T_1(v_1)|_{i-1}$ . Duplicator allows Spoiler a free move of  $v_1$ . Let  $\sigma$  be the  $\Gamma_{i-1}$  class for  $T_1(v_1)|_{i-1}$ . In  $T_2$  Duplicator finds a child  $v_2$  of the root  $R_2$  in  $T_2$  such that  $T_2(v_2)|_{i-1} \in \sigma$ . Duplicator now moves  $v_2$  and then, by induction hypothesis, finds the appropriate response  $w_2 \in T_2(v_2)|_{i-1}$  corresponding to  $w_1$ . For any further moves by the Spoiler with the same  $v_1$  or  $v_2$ , Duplicator plays, inductively, on the two subtrees  $T_1(v_1)|_{i-1}, T_2(v_2)|_{i-1}$ . And if Spoiler chooses some  $y_1 \in B_{T_1}(R_1; i) - T_1(v_1)|_{i-1}$ , then again we repeat the same procedure as above. There are only  $k$  moves, hence Duplicator can continue in this manner and so wins the Ehrenfeucht game.  $\square$

When  $\sigma \in \Gamma_i$  we write  $\Pr[\sigma], \Pr^*[\sigma]$  for the probabilities, in  $T_\lambda, T_\lambda^*$  respectively, that  $\overline{B_T(R; i)}$  is in equivalence class  $\sigma$ . Let  $\Gamma = \Gamma_s$  with  $s = 3^{k+1}$ .

For any first order  $A$  with quantifier depth  $k$  let  $J_A$  be as in (1.17). Applying Theorem 4.2 for each  $i \in J_A$  the class  $\mathcal{B}_i$  splits into finitely many classes  $\tau \in \Gamma$ . Let  $K_A$  denote the set of such classes. The equation (1.18) can be rewritten as

$$\Pr^*[A] = \sum_{\tau \in K_A} \Pr^*[\tau]. \quad (4.4)$$

For  $0 \leq i < k$  set

$$P_i(x) = \Pr[Po(x) = i] = e^{-x} \frac{x^i}{i!}, \quad (4.5)$$

and set

$$P_\omega(x) = \Pr[Po(x) \geq k] = 1 - \sum_{i=0}^{k-1} P_i(x). \quad (4.6)$$

We now make use of a special property of the *Poisson* distribution. Let  $\Omega = \{1, \dots, n\}$  be some finite state space. Let  $p_i \geq 0$  with  $\sum_{i=1}^n p_i = 1$  be some distribution over  $\Omega$ . Suppose  $v$  has *Poisson* mean  $\lambda$  children and each child independently is in state  $i$  with probability  $p_i$ . The distribution of the number of children of each type is the same as if for each  $i \in \Omega$  there were *Poisson* mean  $p_i \lambda$  children of type  $i$  and these values were mutually independent. For example, assumming boys and girls equally probable, having *Poisson* mean 5 children is the same as having *Poisson* mean 2.5 boys and, independently, having *Poisson* mean 2.5 girls.

The probability, in  $T_\lambda$ , that the root has  $u$  children (including  $u = \omega$ ) is then  $P_u(\lambda)$ . Suppose, by induction, that  $P_\tau(x)$  has been defined for all  $\tau \in \Gamma_i$  such that  $\Pr(\tau) = P_\tau(\lambda)$ . Let  $\sigma \in \Gamma_{i+1}$  so that  $\sigma$  is a function  $g : \Gamma_i \rightarrow C$ . In  $T_\lambda$  the root has *Poisson* mean  $\lambda$  children and, for each  $\tau \in \Gamma_i$ , the  $i$ -generation tree rooted at a child is in the class  $\tau$  with probability  $P_\tau(\lambda)$ . By the special property above we equivalently say that the root has *Poisson* mean  $\lambda P_\tau(\lambda)$  children of type  $\tau$  for each  $\tau \in \Gamma_i$  and that these numbers are *mutually independent*. The probability  $P_\sigma(\lambda)$  is then the product, over  $\tau \in \Gamma_i$ , of the probability that a *Poisson* mean  $\lambda P_\tau(\lambda)$  has value  $g(\tau)$ . Setting

$$P_\sigma(x) = \prod_{\tau} P_{g(\tau)}(x P_\tau(x)), \quad (4.7)$$

we have

$$\Pr[\sigma] = P_\sigma(\lambda). \quad (4.8)$$

**Example 4.3.** Continuing Example 4.1, set  $x_i = e^{-\lambda} \lambda^i / i!$  for  $0 \leq i < 4$  and  $x_\omega = 1 - \sum_{i=0}^3 x_i$ . The root has no child with three children with probability  $\exp[-x_3 \lambda]$ . It has one child with at least four children with probability  $\exp[-x_\omega \lambda](x_\omega \lambda)$ . It has at least four children with one child with probability  $1 - \exp[-x_1 \lambda](1 + (x_1 \lambda) + (x_1 \lambda)^2 / 2 + (x_1 \lambda)^3 / 6)$ . It has two children with no children with probability  $\exp[-x_0 \lambda](x_0 \lambda)^2 / 2$ . It has three children with two children with probability  $\exp[-x_2 \lambda](x_2 \lambda)^3 / 6$ . The probability of the event is then the product of these five values.

While Equation (4.8) gives a very full description of the possible  $\Pr[\sigma]$  the following less precise description may be more comprehensible.

**Definition 4.4.** Let  $\mathcal{F}$  be the minimal family of function  $f(\lambda)$  such that

- (i)  $\mathcal{F}$  contains the identity function  $f(\lambda) = \lambda$  and the constant functions  $f_q(\lambda) = q, q \in \mathbb{Q}$ .
- (ii)  $\mathcal{F}$  is closed under finite addition, subtraction and multiplication.
- (iii)  $\mathcal{F}$  is closed under base  $e$  exponentiation. That is, if  $f(\lambda) \in \mathcal{F}$  then  $e^{f(\lambda)} \in \mathcal{F}$ .

We call a function  $f(\lambda)$  *nice* if it belongs to  $\mathcal{F}$ .

In Corollary 4.8 we show that the probability of any first order property, conditioned on the tree being infinite, is actually such a nice function.

**Theorem 4.5.** Then for all  $k$  and all  $i$ , if  $\sigma \in \Gamma_i$  then  $\Pr[\sigma]$  is a nice function of  $\lambda$ .

This is an immediate consequence of the recursion (4.7).

**Example 4.6.** The statement “the root has no children which have no children which have no children” is the union of classes  $\sigma$  with  $k = 1, i = 3$ . It has probability  $\exp[-\lambda \exp[-\lambda \exp[-\lambda]]]$ .

Let  $T_\lambda^{fin}$  denote  $T_\lambda$  conditioned on  $T_\lambda$  being finite. For any  $k, i$  and any  $\sigma \in \Gamma_i$  let  $\Pr^{fin}[\sigma]$  be the probability of event  $\sigma$  in  $T_\lambda^{fin}$ . Assume  $\lambda > 1$ . Let  $p = p(\lambda)$ , the probability  $T_\lambda$  is infinite, be given by (1.1). By duality,  $T_\lambda^{fin}$  has the same distribution as  $T_{q\lambda}$ , where

$$q(\lambda) = 1 - p(\lambda) = \Pr[T_\lambda \text{ is finite}]. \quad (4.9)$$

Thus

$$\Pr^{\text{fin}}[\sigma] = P_\sigma(q\lambda). \quad (4.10)$$

For any  $k, i$  and  $\sigma \in \Gamma_i$

$$\Pr[\sigma] = \Pr^{\text{fin}}[\sigma]q + \Pr^*[\sigma]p \quad (4.11)$$

and hence

$$\Pr^*[\sigma] = p^{-1}[\Pr[\sigma] - \Pr^{\text{fin}}[\sigma]q]. \quad (4.12)$$

For any first order sentence  $A$  of quantifier depth  $k$ , letting  $K_A$  be as in (4.4),

$$\Pr^*[A] = \sum_{\sigma \in K_A} p^{-1}[\Pr[\sigma] - \Pr^{\text{fin}}[\sigma]q]. \quad (4.13)$$

Combining previous results gives a description of possible  $\Pr^*[A]$ .

**Theorem 4.7.** *Let  $A$  be a first order sentence of quantifier depth  $k$ . Let  $K_A$  be as in (4.4) Let*

$$f(x) = \sum_{\sigma \in K_A} P_\sigma(x). \quad (4.14)$$

*Then*

$$\Pr^*[A] = p^{-1}[f(\lambda) - qf(q\lambda)]. \quad (4.15)$$

As before, it is also convenient to give a slightly weaker form.

**Corollary 4.8.** *For any first order sentence  $A$  we may express*

$$\Pr^*[A] = p^{-1}[f(\lambda) - qf(q\lambda)] \quad (4.16)$$

*where  $f$  is a nice function in the sense of Definition 4.4.*

## 5. FURTHER RESULTS

In this paper, we have so far dealt with Galton-Watson trees with *Poisson* offspring distribution. The results of Sections 2 and 3 extend to some other classes of offspring distributions. In this section, we outline briefly these extensions. We consider a general probability distribution  $D$  on  $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ , where  $p_i$  is the probability that a typical node in the random tree has exactly  $i$  children,  $i \in \mathbb{N}_0$ . We shall denote the probabilities under this regime by  $\Pr_D$ . We also assume that the moment generating function of  $D$  exists on a non-degenerate interval  $[0, \gamma]$  on the real line.

Fix an arbitrary finite  $T_0$  of depth  $d_0$ . We assume that  $\Pr_D[T_0] > 0$ . In other words, this means that if  $T$  is the random Galton-Watson tree with offspring distribution  $D$ , then  $\Pr_D[T \cong T_0] > 0$ . Consider the statement

$$A = \{\exists v : T(v) \cong T_0\} \vee \{T \text{ is finite}\}. \quad (5.1)$$

We can show, similar to our results in Section 2, that  $\Pr_D[A] = 1$ , provided (5.6) holds for some  $\alpha \in (0, \gamma]$  and  $0 < \epsilon < 1$ . Of course, the non-trivial case to consider is when  $D$  has expectation greater than 1, as only then does it make sense to talk about the infinite Galton-Watson tree.

The proof of this fact follows the exact same steps as shown in Section 2. We consider again a fictitious continuation  $X_1, X_2, \dots$  which are i.i.d.  $D$ . For every node  $i$ , we let  $I_i$  be the indicator for the event  $\{T(i) \cong T_0\}$ . For a suitable  $\epsilon > 0$  that we choose later, we let

$$Y = \sum_{i=1}^{\lfloor \epsilon^{d_0} s \rfloor} I_i, \quad (5.2)$$

and we define the martingale  $Y_i = E[Y|X_1, \dots, X_i]$  for  $1 \leq i \leq s$ , with  $Y_0 = E[Y]$ . Defining  $g_1$  as in Equation (2.7), we similarly argue that

$$g_1(x) \leq \lfloor x \rfloor + \sum_{i=1}^{\lfloor x \rfloor} X_i. \quad (5.3)$$

The only difference is in the estimation of the probability that  $g_1(\epsilon x)$  exceeds  $x$ . We employ Chernoff bounds again, but we no longer have the succinct form of the moment generating function as in the case of *Poisson*. For any  $0 < \alpha \leq \gamma$ ,

$$\begin{aligned} \Pr[g_1(\epsilon x) > x] &= \Pr[e^{\alpha g_1(\epsilon x)} > e^{\alpha x}] \\ &\leq E[e^{\alpha g_1(\epsilon x)}] e^{-\alpha x} \\ &\leq E[e^{\alpha(\epsilon x + \sum_{i=1}^{\lfloor \epsilon x \rfloor} X_i)}] e^{-\alpha x} \\ &= e^{\alpha \epsilon x} \prod_{i=1}^{\lfloor \epsilon x \rfloor} E[e^{\alpha X_i}] e^{-\alpha x} \\ &= \varphi(\alpha)^{\lfloor \epsilon x \rfloor} e^{-\alpha(1-\epsilon)x}, \end{aligned} \quad (5.4)$$

where  $\varphi(\alpha) = E[e^{\alpha X_1}]$ . Since  $X_1$  is non-negative valued,  $\varphi(\alpha) > 1$  for  $\alpha > 0$ , hence we can bound the expression in (5.4) above by

$$\varphi(\alpha)^{\epsilon x} e^{-\alpha(1-\epsilon)x} = \left\{ \varphi(\alpha)^{\epsilon} e^{-\alpha(1-\epsilon)} \right\}^x. \quad (5.5)$$

If we are able to choose  $\alpha > 0$  such that for some  $0 < \epsilon < 1$ , we have

$$\varphi(\alpha)^{\epsilon} e^{-\alpha(1-\epsilon)} < 1, \quad (5.6)$$

then the exact same argument as in Section 2 goes through, and we have the desired result.

In particular, it is easy to see that (5.6) is indeed satisfied when  $D$  is a probability distribution on a finite state space  $\subseteq \mathbb{N}_0$ .

The sufficient conditions for a tree to be universal nowhere uses the offspring distribution. Once the results of Section 2 hold for a given  $D$ , it is not too difficult to see that the conclusion of Remark 1.10 should hold in this regime as well. We hope to return to this more general setting in our future work.

A further object of future study is a more detailed analysis of  $T_\lambda$  at the critical value  $\lambda = 1$ . While  $\Pr^*$  is technically not defined at the critical value, there may well be some approaches via the insipient infinite tree.

## REFERENCES

- Noga Alon and Joel H. Spencer. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, third edition, 2008. ISBN 978-0-470-17020-5. doi: 10.1002/9780470277331. URL <http://dx.doi.org/10.1002/9780470277331>. With an appendix on the life and work of Paul Erdős.
- Joel Spencer. *The strange logic of random graphs*, volume 22 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 2001. ISBN 3-540-41654-4. doi: 10.1007/978-3-662-04538-1. URL <http://dx.doi.org/10.1007/978-3-662-04538-1>.
- Joel Spencer and Lubos Thoma. On the limit values of probabilities for the first order properties of graphs. In *Contemporary trends in discrete mathematics (Štířín Castle, 1997)*, volume 49 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 317–336. Amer. Math. Soc., Providence, RI, 1999.



Remco van der Hofstad. *Random Graphs and Complex Networks, Vol 1.* Lecture Notes, <http://www.win.tue.nl/~rhofstad/NotesRGCN.pdf>, 2015.

MOUMANTI PODDER, COURANT INSTITUTE OF MATHEMATICAL SCIENCES, NEW YORK UNIVERSITY, 251 MERCER STREET, NEW YORK, NY 10012, UNITED STATES.

*E-mail address:* `mp3460@nyu.edu`.

JOEL SPENCER, PROFESSOR, COMPUTER SCIENCE AND MATHEMATICS DEPTS, COURANT INSTITUTE OF MATHEMATICAL SCIENCES, NEW YORK UNIVERSITY, 251 MERCER STREET, NEW YORK, NY 10012, UNITED STATES.

*E-mail address:* `spencer@cims.nyu.edu`.